John von Neumann Institute for Computing

The Large and Small Scale Structure of the Universe

Stefan Gottlöber, Gustavo Yepes, and Matthias Hoeft

published in

NIC Symposium 2006 ,
G. Münster, D. Wolf, M. Kremer (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **32**, ISBN 3-00-017351-X, pp. 31-38, 2006.

The Large and Small Scale Structure of the Universe

Stefan Gottlöber[1], **Gustavo Yepes**[2], **and Matthias Hoeft**[3]

[1] Astrophysical Institute Potsdam
An der Sternwarte, 14482 Potsdam, Germany
E-mail: sgottloeber@aip.de

[2] Grupo de Astrofísica, Universidad Autónoma de Madrid, Madrid 28049, Spain
E-mail: gustavo.yepes@uam.es

[3] International University Bremen, Campus Ring 1, 28759 Bremen, Germany
E-mail: m.hoeft@iu-bremen.de

We use high resolution cosmological simulations to study the formation and evolution of small and large scale cosmological structures.

1 Introduction

The exciting observational developments of the past couple of decades have been followed closely by comparable progress in our theoretical understanding of the main processes that govern the evolution of structure in the Universe. A substantial part of this progress is due to the increasing possibilities to simulate the formation and evolution of structure on different scales using the new generation of massively parallel supercomputers. The standard model of cosmological structure formation is based on surprisingly few parameters which can be measured at present with high accuracy: the current rate of universal expansion, H_0, the mass density parameter, $\Omega_{\rm mat}$, the value of the cosmological constant, Ω_Λ, the primordial baryon abundance, Ω_b, and the overall normalization of the power spectrum of initial density fluctuations, typically characterized by σ_8, the present-day rms mass fluctuations on spheres of radius $8h^{-1}$ Mpc.

Within the last decade new satellite and earth based observations have been used to determine the cosmological parameters. Most of the cosmologists agree in the "concordance" cosmological model according to which at present the evolution of the Universe is dominated by a cosmological constant ($\Omega_\Lambda = 0.7$). The Universe is spatially flat ($\Omega_\Lambda + \Omega_{\rm mat} = 1$) and the matter consists mainly of dark matter particles (85 %) the nature of which is not yet known. Baryons contribute only by 15 % to the matter density in the Universe. The present expansion rate of the Universe, the Hubble parameter is $H_0 = 70$ km/s/Mpc and the amplitude of the power spectrum is given by $\sigma_8 = 0.9$.

Based on those few numbers numerical simulations allow us to compute the abundance and distribution of galaxies in the Universe. Whereas for the largest objects, clusters of galaxies, simulations and observational surveys do agree very well, there is a discrepancy on the scale of small objects, dwarf galaxies. In fact simulations over-predict substantially the abundance for both, isolated dwarf galaxies in cosmological low density regions and galactic satellites. Numerous solutions have already been proposed, ranging from modifying the dark matter properties to attributing it to the effect of stellar feedback. However, the nature of the discrepancy is still controversial. The abundance of observable dwarf galaxies depends one the large scale environment as well as on local processes. Therefore, a convincing solution can only be found in numerical simulations which take into account both effects.

Our simulations on the NIC supercomputers were in particular designed to derive realistic abundances of small objects in the Universe, like dwarf galaxies. This entails that a cosmological representative volume has to be simulated. The formation of a galaxy is still affected by tidal fields of massive objects at distances of millions of light years. The simulation has to cover such a volume. A dwarf galaxy is in contrast stretched over a few thousand light years at most. The numerical challenge is to cope with more than seven orders of magnitude in spatial and mass resolution. During the past two years we have carried out several simulations which gave considerable new insight into the interplay between large-scale environment and the small scale galaxy formation.

2 Numerical Simulations

During the early inflationary phase of the evolution of the Universe quantum fluctuations became classical perturbations in the density field. These perturbations are predicted to be scale free. At present we cannot measure the fluctuations directly, but we observe their imprint on the cosmic microwave background radiation at recombination, about 400,000 years after the Big Bang. Those temperature fluctuations of the radiation field and the associated density fluctuations are of the order of 10^{-5}. They can be well described by linear perturbation theory. After recombination baryonic density fluctuations are decoupled from the radiation field and start to grow in the potential wells of the dark matter fluctuation. The density fluctuations become soon nonlinear, so that the further evolution can be studied only numerically. In the early eighties first codes have been developed to handle the nonlinear evolution of density perturbations. These codes could follow only 32^3 particles, present codes can handle billions of particles.

Since 85 % of (dark) matter is mainly responsible for the gravitational dynamics, first codes have taken into account only the gravitational interaction of dark matter particles. Nevertheless, these dark matter codes gave and still give many new insights into the formation of large and small scale structures. State-of-the-art hydrodynamical simulations follow also the baryonic gas physics (including radiative cooling) and the formation of stars and their back reaction onto the baryonic gas. Consequently, they are numerically much more expensive and they do not yet reach the same high resolution as pure dark matter codes do.

2.1 The Initial Conditions

The initial Gaussian density fluctuations predicted by inflationary models can be described completely by a power spectrum with a spectral index $n = 1$. This has been confirmed by observations within the present accuracy of measurements. Since fluctuations grow differently during the radiation and the matter dominated phases of the cosmic expansion, the moment when matter density equals radiation density is imprinted on the original power spectrum. The characteristic scale (the horizon at this time) corresponds to the maximum of the present day power spectrum at about 500 Mpc. Moreover, baryonic and dark matter fluctuations grow differently. The resulting changes are described by the transfer functions, which can be calculated by solving numerically the Boltzmann equation. The transfer function depends on the cosmological model, in particular on the measured mean density of dark matter and baryons, the geometry (i.e. the total density) and the expansion rate. The transfer function of the concordance model is well known and the power spectrum of the initial density perturbations can be calculated. The amplitude of the power spectrum is chosen in accordance with the observed temperature fluctuations of the cosmic microwave background and the present day large scale (quasi linear) density fluctuations.

To start a numerical simulation one has first to decide about the size of the box the evolution of which should be simulated. For a given number of particles (which is limited by the power of the computer) this is always a compromise between higher mass resolution (smaller boxes) and a representative volume of the Universe (larger box). Due to the periodical boundary conditions the Universe is assumed to be homogeneous on scales larger than the box size. For a given box size and number of particles one creates a statistical realization of the power spectrum. The displacement of the particles is determined by the entire set of waves that can be represented numerically in the simulation box of size L, i.e. the initial displacements and velocities of N particles are calculated using all waves ranging from the fundamental mode $k = 2\pi/L$ to the Nyquist frequency $k_{\mathrm{Ny}} = 2\pi/L \times N^{1/3}/2$.

If we are interested in the evolution of certain objects in a cosmological environment we use a mass refinement technique. To construct suitable initial conditions, we first create an unconstrained random realization at very high resolution which is limited by the available memory (At the SP4 of NIC we can calculate the initial displacement of up to 2048^3 (~ 8.6 billion) particles.) Then we reduce the resolution by merging particles and assigning to them a velocity and a displacement equal to the average values of the original small-mass particles. We run at small computational costs a simulation which evolves 128^3 or 256^3 particles until the present epoch and selected from this simulation regions or objects in which we are interested.

Now we can construct for this region the initial density field with very high accuracy whereas the surrounding field is given by much larger masses. To this end we go back to the original sample of small-mass particles in the regions of interest. We construct a series of shells around this region where we progressively merge more and more of the particles until the low resolution is reached again far away from the objects of interest. This procedure ensures that the selected objects evolve in the proper cosmological environment and with the right gravitational tidal fields.

2.2 The Code

For most of our simulations we use the publicly available code GADGET-2. It allows to include both the dominant dark matter component and the baryonic matter, i.e. gas and stars. This code was developed by Volker Springel. It is a full MPI N-body code based on the *TREE-PM* algorithm. The gravitational forces acting on each particle are splitted in two parts, a long range force due to distant particles and the short range gravitational forces due to the neighboring particles. The long range force is evaluated by means of the classical *Particle-Mesh* algorithm, which consists of solving the Poisson equation on a regular mesh by Fast Fourier Transforms. The short range forces are computed by means of a Tree algorithm which categorize the particles according to the distance to any other one.

The baryonic component is also discretized in particles carrying the fluid information with them. The hydrodynamical quantities at any position in space can be found using interpolation from the fluid particles. The equations of gas dynamics are solved by means of the *Smoothed Particle Hydrodynamics* method.

In addition to the modeling of gravity and gas dynamics, the code also implements additional physics for the baryons, like Compton cooling due to interactions with photons of cosmic background radiation, radiative cooling due to atomic recombination and photoionization by ultraviolet photons due to the presence of a homogeneous background coming from quasar and galactic sources. When gas cools, its density increases and eventually it will form dense molecular clouds in which stars will be born. Although the physics of this process is not well known, the code implements a model of the star-gas interactions that attempts to mimic the multiphase nature of the interstellar medium. The final outcome of the model is to transform gas particles into collisionless star particles representing star-burst.

Switching on and off the different models implemented in the code, one can use it to simulate the pure dark matter gravitational evolution (using the N-body algorithm only), or the dark matter and gas dynamics (using N-body and SPH), or everything together (N-body + SPH + star formation).

All the different parts of the code are done in parallel using MPI routines. The parallelization is done by domain decomposition of the simulation volume to distribute the particles among the different processors. If only N-body and SPH algorithms are used, the code scales pretty well with the number of processors. When the non-adiabatic physics module (cooling, heating, star formation) is switched on, the enormous differences in density and timing between star forming regions and the rest of the simulation makes the scaling not very efficient beyond several dozens of processors.

2.3 The Data Analysis

With 1 billion of particles a numerical code produces per time step at least 28 Gb of data (3 positions, 3 velocities and one id for each particle), in case of multi mass simulations also masses must be stored. It is a challenge to find structures and substructures in the distribution of those particles. To find structures at virial over-density one can use a friends-of-friends algorithm with a certain linking length (0.17 of the mean interparticle distance for the concordance model at redshift zero). The resulting particle clusters are in general three-axial objects. Substructures can be identified as particle clusters at smaller linking lengths (higher over-densities). The more different linking lengths are used the better substructures will be resolved. Thus a whole hierarchy of friends-of-friends clusters have to be calculated. To this end we have developed a hierarchical friends-of-friends algorithm which is based on the calculation of the minimum spanning tree of the given particle distribution.

The minimum spanning tree of any point distribution is a unique, well defined quantity which describes the clustering properties of the point process completely. The minimum spanning tree of n points contains $N - 1$ connections. Based on the minimum spanning tree we sort the particles in such a way that we get a cluster-ordered sequence $P = \{p_1, p_2, ..., p_n\}$. Any cluster is a segment of the sequence P, i.e. it consist of points $p_i, p_{i+1}, ..., p_j$ for some indexes i and j. Neighboring clusters, i.e. clusters which merged immediately after increasing r, are neighboring segments on P. Let us denote the length at which clusters $p_i, p_{i+1}, ..., p_j$ and $p_{j+1}, p_{j+2}, ..., p_k$ merge, by $r_{j+1/2}$. The sequences

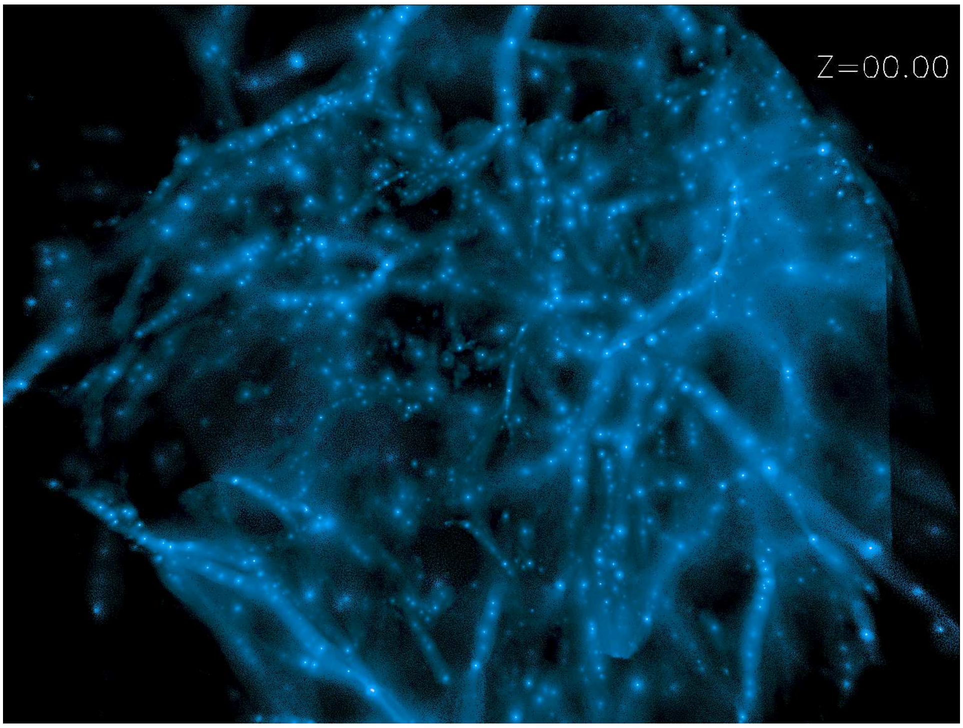

Figure 1. Gas distribution in a void region of $10h^{-1}$ Mpc in diameter. The color represent the gas density. The bright spots are the areas of high density gas that has cooled down to form stars in dwarf galaxies.The 3d positions of 2 million gas particles of mass $10^5 M_\odot$ are shown.

P and R are sufficient for deriving the complete list of clusters at any linking length r. In fact, the segment $p_i, p_{i+1}, ..., p_j$ of the sequence P is an r-cluster if and only if $r_{i-1/2} > r$, $r_{j+1/2} > r$ and $r_{k+1/2} \leq r, \quad k = i, i+1, ..., j-1$. In other words, if all points would be located on a line with distances $r_{j+1/2}, j = 1, 2, ..., n-1$ between neighboring points, the line would break into the sequence of all r-clusters after cutting of all segments larger than r. Obviously, the sequences P and R (each of length $n_p \times 4$ byte) is the most compact form to store the information about the whole hierarchy of friends-of-friends clusters.

The minimum spanning tree and the cluster analysis are done within MPI programs. The basic idea of parallelization is the calculation of separate trees in different regions of the box and merging the subtrees later on. The algorithm is very fast. We need on 8 CPUs of the SP4 at NIC about 30 minutes for the MST and another 10 minutes for the FOF analysis on 11 density (resp. linking length) levels for a simulation with 512^3 particles.

3 Problems at Short Scales: The Overproduction of Dwarf Galaxies

The major problem of the standard Cold Dark Matter (CDM) scenario of structure formation is the overproduction of small structures. All CDM simulations predict that there should be many dwarf galaxies orbiting around the big ones, like our Milky Way, and that

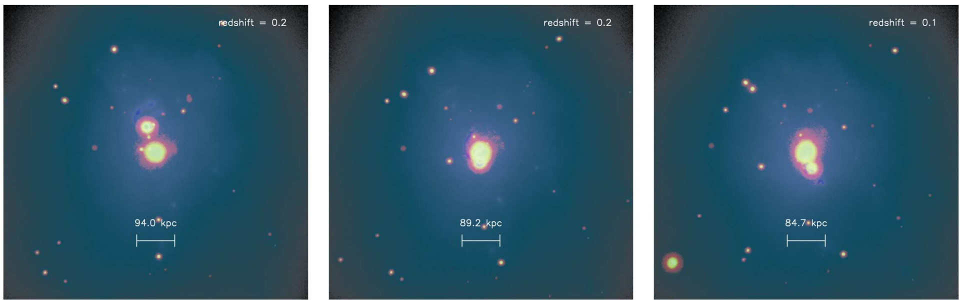

Figure 2. Interaction between two galaxies. Blue gives the gas distribution and yellow/red the distribution of stars. This galaxy is a small part from a $50h^{-1}$Mpc simulation box. One can also clearly see the distribution of satellite galaxies and gas clouds.

there should be a large population of such objects in void regions, in which no bright galaxy has been formed. In the Universe, neither we see so many satellites around disk galaxies nor we detect a population of dwarfs in voids. One way out of this discontenting situation is the claim that baryonic physics is responsible for the absence of luminous matter within the small dark halos. After all, the predictions from N-body simulations correspond to the dark matter distribution only. The baryons could have a different distribution than the dark component and stars could be formed only under special conditions.

In order to get a deeper insight into this question we have carried out a series of hydrodynamical simulations with increasing resolution of a void region of 10 Mpc in diameter, extracted from a larger box. We have included all the relevant physical processes of the baryonic matter: i.e cooling, UV photoionization, star formation and stellar feedback.

In Figure 1 we show the gas density within the void. The little white dots indicate where the gas has collapse due to cooling and have formed a dwarf galaxy. This plot shows that a void region is far from being empty. Moreover, it mimics the filamentary structure that is seen at much larger scales. But how many of these little gas spots have been able to produce a reasonable amount of stars that can allow us to see them today with powerful enough telescopes? Thanks to the kind of simulations we have done in the JUMP supercomputer at NIC we can compute the luminosity of these objects as a function of the physics we put in. The main conclusion of our study is that, although there is a considerable effect of suppressing star formation in these little objects due to the cosmic ultraviolet photoionization background, it is not enough to actually suppress all dwarf galaxies to form stars. If results from these simulations were correct, we should see a substantial amount of faint objects in voids. Therefore, another mechanism apart from photoionization should be responsible for suppressing the luminosity of dwarf dark halos in voids.

We are currently investigating this problem of excess of short scale structure in regions of higher density. To this end, we are now simulating, with the same resolution as in the void regions, other parts of the original box in which denser structures form, ranging from isolated galaxies, small groups and even clusters of galaxies. These simulations are of course much more demanding in terms of computational resources and will take longer to finish. In Figure 2 we show one example of the interaction of two galaxies and their satellites.

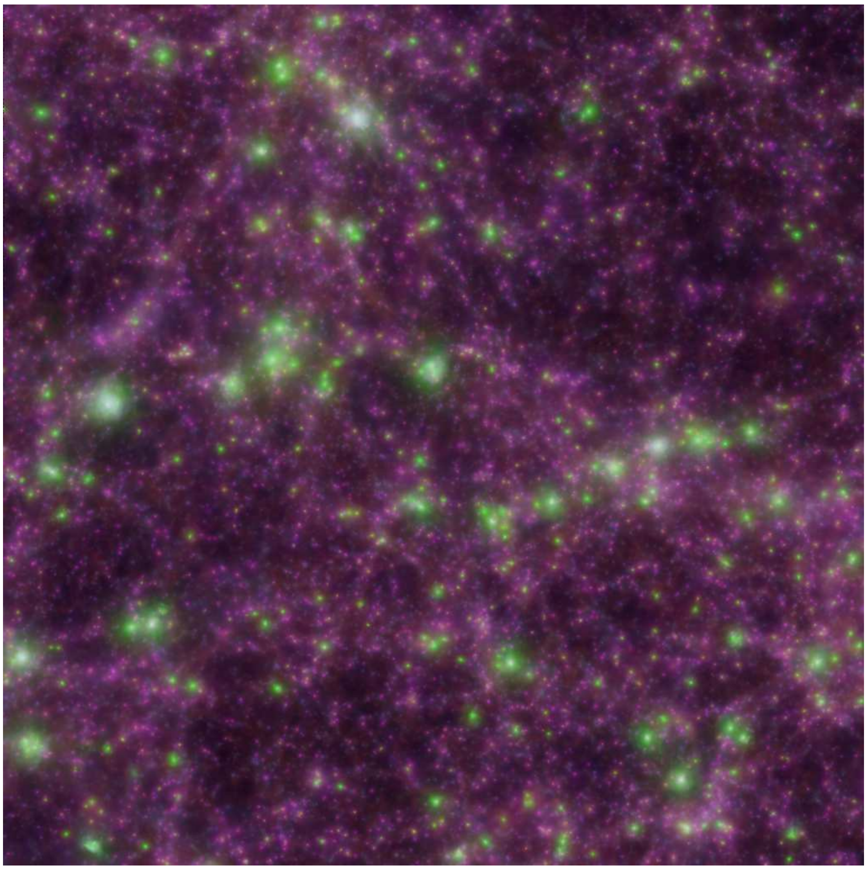

Figure 3. Gas distribution, color coded according to temperature, for a small part ($75h^{-1}$ Mpc) of a hydrodynamical simulation with 2 billion particles.

4 Large Scale Structure of the Universe

The problem of the formation of Large Scale Structures in the Universe is another interesting research topic that can only be investigated with very big simulations. If only collisionless dark matter is used in a simulation, the number of particles that can currently be treated is of the order of 10^{10}, using computational resources equivalent to 16 nodes of JUMP. Although these simulations provide a very detailed description of the structure formed by dark matter, they lack the baryonic physics, so a direct comparison with observational results is not possible. If one wants to include gas dynamics in the simulation, then one has to reduce the number of particles. We have investigated the structures formed both in gas and dark matter by means of these kind of large simulations. In particular we studied the correlation of orientations of clusters of galaxies and of super-clusters using a simulation with 2×512^3 particles done in JUMP.

Using 16 nodes of 32 CPUs each on JUMP we can run a 2 billion simulation with dark matter and gas together. Recently, we have investigated the structures formed both in gas and dark matter using such an extreme simulation. Due to the large volume ($500h^{-1}$ Mpc cubic box) most of the structures in the simulation evolve independently to some extent. Thus we could resort to the slower communication of a PC cluster system and run the simulation in the new massively parallel IBM supercomputer *MareNostrum* (BSC, Spain) during about 500 wall clock hours using 512 PowerPC processors (256000 CPU hours, 1 Tbyte of memory). All the post-processing of the data has been done again in JUMP, where we can use our memory-intensive MPI + OpenMP programs for analysis. In Figure 3 the gas distribution in a small part of the simulation is shown.

5 Outlook

Simulations with a very large number of particles running in parallel supercomputers are the main tool for studying cosmological structure formation. But running a big simulation is just the starting point of the research work. All the data processing takes usually as much computing time as the simulation itself.

Over the next years our groups are planning to simulate as accurate as never before the formation and evolution of our local neighborhood in the Universe. These constrained simulations will be done within one of the DEISA Extreme Computing projects in which NIC participates.

Acknowledgments

Simulations described here would be impossible without the excellent user support and help by NIC and other supercomputer centers.

References

1. S. Basilakos, M. Plionis, G. Yepes, S. Gottlöber, V. Turchaninov, *The Shape-Alignment relation in* Λ*CDM Cosmic Structures*, MNRAS, accepted (astro-ph/0505620).
2. S. Gottlöber, V. Turchaninov, *Halo Shape and its Relation to Environment* EDP Sciences (Paris), in print, astro-ph/0511675.
3. M. Hoeft, G. Yepes, S. Gottlöber, V. Springel, *Dwarf galaxies in voids: Suppressing star formation with photo-heating*, MNRAS, submitted, astro-ph/0501304.
4. V. Springel, *The cosmological simulation code* GADGET-2, MNRAS 364 (2005), 1105.